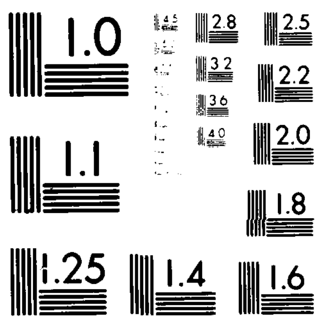
AD-A097 531 NAVAL RESEARCH LAB WASHINGTON DC F/G 9/4
MINIMUM-CROSS-ENTROPY SPECTRAL ANALYSIS OF MULTIPLE SIGNALS.(U)
APR 81 R W JOHNSON, J E SHORE
UNCLASSIFIED NRL-MR-4492 NL

END
DATE
FILMED
5-81
DTIC


1.0
1.1
1.25
1.4
1.6
1.8
2.0
2.2
2.5
2.8
3.2
3.6
4.0

SECURITY CLASSIFICATION OF THIS PAGE (When Data Entered)

REPORT DOCUMENTATION PAGE

READ INSTRUCTIONS BEFORE COMPLETING FORM

1. REPORT NUMBER: NRL Memorandum Report 4492 — 2. GOVT ACCESSION NO.: AD-A097 531	3. RECIPIENT'S CATALOG NUMBER
4. TITLE (and Subtitle): MINIMUM-CROSS-ENTROPY SPECTRAL ANALYSIS OF MULTIPLE SIGNALS	5. TYPE OF REPORT & PERIOD COVERED: Interim report on a continuing NRL problem.
	6. PERFORMING ORG. REPORT NUMBER
7. AUTHOR(s): R. W. Johnson [redacted] J. E. Shore	8. CONTRACT OR GRANT NUMBER(s)
9. PERFORMING ORGANIZATION NAME AND ADDRESS: Naval Research Laboratory, Washington, DC 20375	10. PROGRAM ELEMENT, PROJECT, TASK AREA & WORK UNIT NUMBERS: 61153N; RR021-05-42; 75-0107-0-1
11. CONTROLLING OFFICE NAME AND ADDRESS: NRL-MR-4492	12. REPORT DATE: Apr[redacted] [redacted]81
	13. NUMBER OF PAGES: 34
14. MONITORING AGENCY NAME & ADDRESS (if different from Controlling Office)	15. SECURITY CLASS. (of this report): UNCLASSIFIED
	15a. DECLASSIFICATION/DOWNGRADING SCHEDULE

16. DISTRIBUTION STATEMENT (of this Report)

Approved for public release; distribution unlimited.

17. DISTRIBUTION STATEMENT (of the abstract entered in Block 20, if different from Report)

18. SUPPLEMENTARY NOTES

19. KEY WORDS (Continue on reverse side if necessary and identify by block number)

Spectral anlaysis	Information theory
Maximum entropy	Discrimination information
Minimum cross entropy	Itakura-Saito distortion
Directed divergence	Noise suppression

20. ABSTRACT (Continue on reverse side if necessary and identify by block number)

This paper presents a new information-theoretic method for simultaneously estimating a number of power spectra when a prior estimate of each is available and new information is obtained in the form of values of the autocorrelation function of their sum. A derivation of the method from the principle of minimum cross entropy is given, and the method is compared with maximum-entropy spectral analysis and with minimum-cross-entropy spectral analysis, of which it is a generalization. Some basic mathematical

(Continues)

DD FORM 1 JAN 73 1473 EDITION OF 1 NOV 65 IS OBSOLETE
S/N 0102-014-6601

SECURITY CLASSIFICATION OF THIS PAGE (When Data Entered)

251700 LB

20. ABSTRACT (Continued)

properties are discussed. Two illustrative numerical examples are included, one based on synthetic spectra, and one based on actual speech data.

CONTENTS

MINIMUM-CROSS-ENTROPY SPECTRAL ANALYSIS OF MULTIPLE SIGNALS

I. INTRODUCTION AND BACKGROUND

We present here an information-theoretic method for simultaneously estimating a number of power spectra when a prior estimate of each is available and new information is obtained in the form of values of the autocorrelation function of their sum. The method applies for instance when one obtains autocorrelation measurements for a signal with independent additive interference, and one has some prior knowledge concerning the signal and the noise spectra; the result is signal- and noise-spectrum estimates that take both the prior estimates and the autocorrelation information into account. One thus obtains a procedure for noise suppression that offers some advantages over more traditional procedures, such as those based on spectral subtraction.

The present method is a generalization of minimum-cross-entropy spectral analysis [1], which is in turn a generalization of maximum-entropy (or linear-predictive or autoregressive) spectral analysis [2], [3]. All these methods proceed from autocorrelation values. Minimum-cross-entropy spectral analysis (MCESA) differs from maximum-entropy spectral analysis (MESA) in that it explicitly uses a prior estimate of the power spectrum; it reduces to MESA as a special case when the prior estimate is uniform and one of the given autocorrelation values is for zero lag. The present method, multi-signal MCESA, differs from MCESA in that it treats an arbitrary number of independent

Manuscript submitted February 17, 1981.

spectra simultaneously; in the special case of a single spectrum, it becomes identical to MCESA.

MESA may be regarded as an application of the principle of maximum entropy [4], [5]; single- and multi-signal MCESA are applications of a generalization of that principle, the principle of minimum cross entropy (also called minimum discrimination information, directed divergence, I-divergence, relative entropy, or Kullback-Leibler number) [6], [7], [8], [9], [10], [11]. In the remainder of this section, we describe these spectrum-analysis methods further and include some background on the principle of minimum cross entropy. Section II contains a derivation of our multiple-signal estimator, and section III discusses a few of its general properties. Section IV presents two numerical examples, one of which is based on measured samples of speech signals and noise. Finally, section V contains a concluding discussion.

A. MESA and MCESA

MESA addresses the following problem: *estimate the power spectrum* S(f) *of a real, band-limited, stationary process, given values of the autocorrelation function*

$$R(t) = 2\int_0^W df\, S(f)\cos 2\pi ft$$

for finitely many lags $t = t_r$, $r = 0,\ldots, M$. (Here W is the bandwidth.) The solution proposed by Burg [2], [3] is to choose the estimate Q of S that maximizes

$$\int_0^W df \log Q(f) \tag{1}$$

subject to the constraint that the autocorrelation function assume the given values:

$$R(t_r) = 2\int_0^W df\ Q(f)\cos 2\pi f t_r \tag{2}$$

The resulting estimator has the form

$$Q(f) = \frac{1}{\sum_r 2\beta_r \cos 2\pi f t_r}, \tag{3}$$

where the coefficients β_r $(r = 0,\ldots,M)$ are chosen so that Q satisfies (2).

MCESA is applicable to the problem of estimating S(f) when, in addition to the autocorrelation values, a <u>prior</u> estimate P of S is given; P may be thought of as the best guess at S we could make in the absence of autocorrelation data. The MCESA estimator has the form [1]

$$Q(f) = \frac{1}{1/P(f) + \sum_r 2\beta_r \cos 2\pi f t_r}, \tag{4}$$

where again the β_r are chosen so that Q satisfies the constraints (2). We call Q the <u>posterior</u> estimate of S based on the prior estimate P and constraints (2). This estimator can be obtained directly from the minimum-cross-entropy principle [1]; it can also be obtained by minimizing the Itakura-Saito distortion measure [12]

$$\int df \left(\frac{Q(f)}{P(f)} - \log \frac{Q(f)}{P(f)} - 1 \right)$$

subject to (2) [1]. When P(f) is uniform, and one of the autocorrelation values is at lag zero (say $t_0 = 0$), we can write (4) in the form (3), since the constant 1/P can be absorbed into the coefficient β_0. Thus in this case MCESA reduces to MESA.

For multi-signal MCESA, the problem is to estimate the power spectra $S_i(f)$ of a number of independent processes, given values of the total autocorrelation

$$R(t) = 2 \sum_i \int_0^W df\ S_i(f) \cos 2\pi f t$$

and a prior estimate P_i for each S_i. The estimator has the form

$$Q_i(f) = \frac{1}{1/P_i(f) + \sum_r 2\beta_r \cos 2\pi f t_r}, \tag{5}$$

where the β_r are chosen so that the constraint equations

$$R(t_r) = 2 \sum_i \int_0^W df\ Q_i(f) \cos 2\pi f t_r \tag{6}$$

are satisfied. Note that the summation term in the denominator in (5) is independent of i. In Section III we derive the estimates (5) directly from the principle of minimum cross entropy. We also show that they can be obtained by minimizing the sum

$$\sum_i \int df \left(\frac{Q_i(f)}{P_i(f)} - \log \frac{Q_i(f)}{P_i(f)} - 1 \right)$$

of Itakura-Saito distortions subject to the constraints (6). Equations (5) and (6) reduce to (4) and (2) when there is only one spectrum S_i; thus multi-signal MCESA reduces to ordinary MCESA in case there is only one signal.

B. Cross-Entropy Minimization

The principle of minimum cross entropy is a general method for inference about probability distributions when information is avaliable in the form of expectation values of known functions.

Let $q^{\dagger}$ be a probability density on a space of states $\underset{\sim}{x}$ of some system. Suppose that $q^{\dagger}$ is not known, but there is some prior density p (on the same space) that is our current estimate of $q^{\dagger}$. Now suppose we gain new information about $q^{\dagger}$ in the form of expectation values

$$\int d\underset{\sim}{x}\; q^{\dagger}(\underset{\sim}{x}) g_r(\underset{\sim}{x}) = \bar{g}_r \tag{7}$$

of known functions g_r. In general, these constraints do not determine $q^{\dagger}$ uniquely: the equations (7) are satisfied by other densities q than $q^{\dagger}$ (but not necessarily by p). The problem to be solved is, given p and the constraints (7), to make the best possible choice of a new (or posterior) estimate q of $q^{\dagger}$. The principle of minimum cross entropy states that one should choose that density q, among all the densities that satisfy the constraints, that has the least cross entropy

$$H(q,p) = \int d\underset{\sim}{x}\; q(\underset{\sim}{x}) \log(q(\underset{\sim}{x})/p(\underset{\sim}{x})) \tag{8}$$

with respect to p.

Given a positive prior probability density p, if there exists a posterior q that minimizes the cross entropy and satisfies the constraints (7), it has the form

$$q(\underset{\sim}{x}) = p(\underset{\sim}{x}) \exp\left(-\lambda - \sum_r \beta_r g_r(\underset{\sim}{x})\right) \tag{9}$$

with the possible exception of a set of states on which the constraints imply that q vanishes [6, p. 38], [10]. In (9), λ and β_r are Lagrange multipliers whose values are determined by the normalization constraint

$$\int d\underset{\sim}{x}\ q(\underset{\sim}{x}) = 1 \tag{10}$$

and by the constraints (7), respectively. Conversely, if there are values for λ and the β_r for which the constraints are satisfied, then the solution exists and is given by (9) [10]. Conditions for existence of solutions are given by Csiszár [10].

One could imagine using a procedure based on minimization of some function of q and p other than H(q,p). In what sense does minimizing cross entropy yield the best estimate q of $q^{\dagger}$? One answer to this question is provided by recent work [7] that characterizes cross-entropy minimization as an inference procedure by means of certain consistency axioms. In describing this work, it is usful to view an inference procedure as an operator ∘ that takes two arguments, a prior probability density p and new constraint information I of the form (7), and yields a posterior probability density p∘I. It is assumed in [7] that ∘ is implemented by minimization of some well behaved function H'(q,p) -- that is, that q = p∘I is defined as that density, among all the densities that satisfy the constraints I, for which H'(q,p) is least. It is further assumed that the operator ∘ satisfies consistency axioms that, informally, require different ways of taking information I into account (for example, in different coordinate systems) to lead to equivalent results. It is then shown to follow from the assumptions that p∘I equals the result of minimizing the cross entropy H(q,p). The axioms do not imply that H' must be H -- for instance a monotonic function of H would do just as well -- but they do uniquely characterize the result $p \circ I$ of the minimization: cross-entropy minimization is uniquely correct in the sense that minimization of any other functional either gives the same result or leads to a contradiction with one of the axioms.

Other justifications for the use of cross-entropy minimization can be based on cross entropy's properties as an information measure [6], [10], [13], [14]. For instance, $H(q,p)$, informally speaking, measures the distortion, "information dissimilarity," or "information divergence" of q from p. $H(q,p)$ can be interpreted as the amount of information needed to change a prior p into the posterior q or to determine q given p [14]; indeed,

$$H(q,p) = H(q^{\dagger},p) - H(q^{\dagger},q) \tag{11}$$

holds when $q = p \circ I$ is defined by cross-entropy minimization [10], [14] In these terms the minimum-cross-entropy principle is intuitively justified as the choice of posterior q that introduces the least distortion, least additional information, or fewest unjustified assumptions consistent with the given constraints. From (11) it follows that $H(q^{\dagger},q) \leqslant H(q^{\dagger},p)$. Thus the posterior q is closer to $q^{\dagger}$ in the cross-entropy sense than is the prior p.

Yet another justification for cross-entropy minimization is provided by the "expectation-matching" property [14], which states that for an arbitrary fixed density q^* and densities q of the form (9), $H(q^*,q)$ is least when the expectations of q match those of q^*. In particular, it follows that $q = p \circ I$ is not only the density satisfying (7) that minimizes $H(q,p)$, but also the density of the form (9) that minimizes $H(q^{\dagger},q)$. Hence $p \circ I$ is not only closer to $q^{\dagger}$ than is p, but it is the closest possible density of the form (10). The expectation-matching property is a generalization of a property of orthogonal polynomials [15, p.12] that, in the case of speech analysis [16], is called the "correlation-matching property" [17, ch. 2]. For further justifications see [7], [14].

II. DERIVATION

We assume the time-domain signal is a sum of stationary random processes $g_i(t)$, $i = 1,\ldots, K$. In many applications, K will be 2 -- one signal process and one noise process -- but the case of arbitrary K is no harder than K = 2, so we do the derivation in that generality. It is convenient to work with discrete-spectrum approximations to the g_i [1], [18],

$$s_i(t) = \sum_{k=1}^{N} (a_{ik}\cos 2\pi f_k t + b_{ik}\sin 2\pi f_k t) ,$$

where the a_{ik} and the b_{ik} are random variables and the f_k are non-zero frequencies, not necessarily uniformly spaced. We write x_{ik} for the power of the process s_i at frequency f_k,

$$x_{ik} = \frac{1}{2}(a_{ik}^2 + b_{ik}^2) ,$$

and will describe the processes in terms of a joint probability density $q^\dagger(\underset{\approx}{x}) = q^\dagger(\underset{\sim}{x}_1,\ldots, \underset{\sim}{x}_K)$, where $\underset{\sim}{x}_i$ stands for $(x_{i1},\ldots, x_{iN})$. The marginal density for each x_i is defined by

$$q_i^\dagger(\underset{\sim}{x}_i) = \int q^\dagger(\underset{\approx}{x}) \prod_{j\neq i} d\underset{\sim}{x}_j ,$$

where each component x_{jk} of the variables of integration ranges from 0 to ∞.

Let $P_{ik} = P_i(f_k)$ be prior estimates of the power spectra of the s_i. Then we may take

$$p_i(\underset{\sim}{x}_i) = \prod_{k=1}^{N} (1/P_{ik})\exp(-x_{ik}/P_{ik}) \tag{12}$$

as prior estimates of $q_i^{\dagger}(\underset{\sim}{x}_i)$. The assumed exponential form is equivalent to a Gaussian distribution in the amplitude variables a_{ik} and b_{ik}; for justification of its use, see [1], [19]. Note that the coefficients are chosen so that the expected value of the power x_{ik} of the process s_i at frequency f_k is equal to the prior estimate:

$$P_{ik} = \int d\underset{\sim}{x}_i p_i(\underset{\sim}{x}_i)x_{ik} .$$

Since we assume independence of $\underset{\sim}{x}_i$ and $\underset{\sim}{x}_j$ $(i \neq j)$, our prior estimate of $q^{\dagger}$ becomes

$$p(\underset{\approx}{x}) = \prod_{i=1}^{K} \prod_{k=1}^{N} (1/P_{ik})\exp(-x_{ik}/P_{ik}) . \tag{13}$$

The spectral power of each process s_i at frequency f_k is given by

$$T_{ik} = \int d\underset{\approx}{x}\, q^{\dagger}(\underset{\approx}{x})x_{ik} , \tag{14}$$

and the autocorrelation function of each s_i is

$$R_{ir} = \sum_{k=1}^{N} c_{rk}T_{ik} , \tag{15}$$

where $c_{rk} = 2 \cos 2\pi t_r f_k$. Suppose we obtain information about $q^{\dagger}$ in the form of autocorrelation values for the sum of the s_i,

$$R_r = \sum_{i=1}^{K} R_{ir} ,$$

$r = 0,\ldots, M$, where $t_0 = 0$. In view of (14) and (15), this has the form of linear constraints on expectation values of $q^\dagger$:

$$R_r = \int d\underset{\approx}{x}\; q^\dagger(\underset{\approx}{x}) \sum_{i=1}^{K} \sum_{k=1}^{N} c_{rk} x_{ik} \; . \qquad (16)$$

Applying the principle of minimum cross entropy to the prior (13) and constraints (16) yields a posterior estimate q of $q^\dagger$ given by

$$q(\underset{\approx}{x}) = p(\underset{\approx}{x}) \exp\left(- \lambda - \sum_{r=0}^{M} \beta_r \sum_{i=1}^{K} \sum_{k=1}^{N} c_{rk} x_{ik}\right)$$

$$= e^{-\lambda} \prod_{i=1}^{K} \prod_{k=1}^{N} (1/P_{ik}) \exp\left(- x_{ik}/P_{ik} - \sum_{r=0}^{M} \beta_r c_{rk} x_{ik}\right)$$

$$= e^{-\lambda} \prod_{i=1}^{K} \prod_{k=1}^{N} (1/P_{ik}) \exp(-A_{ik} x_{ik}) ,$$

where λ and the β_r are Lagrange multipliers corresponding to the constraints (10) and (16), respectively, and $A_{ik} = 1/P_{ik} + \sum_r \beta_r c_{rk}$. Because of the normalization constraint (10), this becomes

$$q(\underset{\approx}{x}) = \prod_{i=1}^{K} \prod_{k=1}^{N} A_{ik} \exp(-A_{ik} x_{ik}) \; . \qquad (17)$$

The posterior estimate of the power spectrum of s_i is

$$Q_{ik} = \int d\underset{\approx}{x}\; q(\underset{\approx}{x}) x_{ik} = \frac{1}{A_{ik}} \; ;$$

thus

$$Q_{ik} = \frac{1}{1/P_{ik} + \sum_r \beta_r c_{rk}}, \tag{18}$$

where the β_r must be chosen so that the constraints

$$R_r = \sum_{i=1}^{K} \sum_{k=1}^{N} c_{rk} Q_{ik} \tag{19}$$

are satisfied. Equations (18) and (19) are simply discrete analogs of (5) and (6).

When p and q are given by (13) and

$$q(\underset{\approx}{x}) = \prod_{i=1}^{K} \prod_{k=1}^{N} (1/Q_{ik}) \exp(-x_{ik}/Q_{ik})$$

(cf. (17)), the cross entropy (8) can be calculated explicitly:

$$H(q,p) = \sum_{i=1}^{K} \left[\sum_{k=1}^{N} \left(\frac{Q_{ik}}{P_{ik}} - \log \frac{Q_{ik}}{P_{ik}} - 1 \right) \right]. \tag{20}$$

The quantity in brackets is a discrete analog of the Itakura-Saito distortion measure [12], [16] of P_i with respect to Q_i; cross-entropy minimization is thus equivalent to choosing the Q_i so as to minimize the sum of Itakura-Saito distortions. We obtain an alternative derivation of (18) by minimizing the right side of (20) directly, subject to the constraints (19). Namely, we form the expression

$$\sum_{i=1}^{K} \sum_{k=1}^{N} \left(\frac{Q_{ik}}{P_{ik}} - \log \frac{Q_{ik}}{P_{ik}} - 1 \right) + \sum_{r=0}^{M} \beta_r \sum_{i=1}^{K} \sum_{k=1}^{N} c_{rk} Q_{ik}$$

involving Lagrange multipliers β_r, and we set the partial derivative with respect to each Q_{ik} equal to zero:

$$1/P_{ik} - 1/Q_{ik} + \sum_{r=0}^{M} \beta_r c_{rk} = 0 \ .$$

This yields (18).

III. PROPERTIES

In this section we discuss three miscellaneous properties of our multi-signal method. We call the first "order preservation"; briefly, it states that the method preserves the relative magnitudes of the priors. The second, "preservation of independence," is related to the assumption of statistical independence of the processes s_i; it follows from a generalization of the property of cross-entropy minimization that was called "system independence" in [7] and [14]. The third is related to a phenomenon that we call "prior washout" and that occurs when a posterior resulting from one analysis is used as a prior for a subsequent analysis; we compare and contrast the behavior of the single- and multi-signal methods in this situation.

A. Order Preservation

Let P_i and P_j be two prior spectra and Q_i and Q_j be corresponding posterior spectra resulting from a multi-signal MCESA analysis. The order-preservation property is the observation that for each frequency f_k we have $Q_i < Q_j$, $Q_i = Q_j$, or $Q_i > Q_j$ if and only if $P_i < P_j$, $P_i = P_j$, or $P_i > P_j$, respectively. This follows from the form of the representation of the Q_i in (5). The property accords well with intuition:

if we expect a priori that s_i has greater power than s_j at frequency f_k, that expectation should not be altered by new information that concerns only the sum of the two powers.

B. Preservation of Independence

In (13), we wrote the prior probability density p in the form

$$p(\underset{\approx}{x}) = \prod_{i=1}^{K} p_i(\underset{\sim}{x}_i)$$

(cf. (12)) to reflect the initial assumption that the $\underset{\sim}{x}_i$ are independent. Preservation of independence is the property that the posterior density q has the same form,

$$q(\underset{\approx}{x}) = \prod_{i=1}^{K} q_i(\underset{\sim}{x}_i)$$

(cf. (17)), so that the $\underset{\sim}{x}_i$ remain independent after the prior density is replaced by the posterior. This posterior independence would be a simple consequence of the system independence property of [7] and [14] if the constraints (16) were of the form

$$R_r = \int d\underset{\approx}{x}\, q^{\dagger}(\underset{\approx}{x}) g_r(\underset{\sim}{x}_{i(r)})$$

-- that is, if each constraint involved only one of the sets x_i of variables (where which set was involved might depend on the constraint). System independence was one of the consistency axioms in [7]; informally, it states that it doesn't matter whether independent constraint information about separate systems with independent priors is accounted for separately, for each system, or jointly, by treating the system as one composite system. In the present case, the constraints have the more general form

$$R_r = \int d\underset{\approx}{x}\; q^{\dagger}(\underset{\approx}{x}) \sum_{i=1}^{K} g_{ri}(\underset{\sim}{x}_i)$$

-- each constraint involves a linear combination of functions, each involving one of the $\underset{\sim}{x}_i$. Nevertheless, posterior independence still follows from prior independence in this more general case.

C. Prior Washout

The phenomenon we are here calling "prior washout" was mentioned in [14] in connection with "Property 14." Property 14, in slightly specialized form, states the following. Let p be a prior probability density. Let $I^{(1)}$ and $I^{(2)}$ be sets of constraints of the form (7), but with the right side replaced by $\bar{g}_r^{(1)}$ for $I^{(1)}$ and by $\bar{g}_r^{(2)}$ for $I^{(2)}$; that is, $I^{(1)}$ and $I^{(2)}$ both constrain the expectations of the same set of functions g_r, but the expected values may differ. Then, in terms of the $\circ$ operator,

$$(p \circ I^{(1)}) \circ I^{(2)} = p \circ I^{(2)} \;;$$

the effects of taking the information $I^{(1)}$ into account are completely washed out when $I^{(2)}$ is taken into account.

One consequence of prior washout is a similar property of single-signal MCESA. For definiteness, consider a speech-processing system; say we wish to estimate the speech spectra $S^{(1)}$, $S^{(2)}$, ... in a succession of analysis frames, and we can measure the speech autocorrelations $R_r^{(1)}$, $R_r^{(2)}$, ... in these frames at a fixed set of lags r. Starting with a prior spectral estimate P, suppose we form a posterior estimate $Q^{(i)}$ for a frame i by taking the autocorrelation information $R^{(i)}$ into account. Suppose we then use this posterior $Q^{(i)}$ as a prior estimate for a later frame j and obtain a posterior estimate for that frame by taking $R^{(j)}$ into account. Prior washout implies that the result $Q^{(j)}$ is the same that we would have gotten if we had used P instead of $Q^{(i)}$ as the prior estimate for frame j; taking $R^{(j)}$ into account completely washes out the effects of having taken $R^{(i)}$ into account.

This property has implications for certain noise-suppression schemes in which one might envision using MCESA. Suppose that additive noise is present in a speech-analysis system. It is often possible to detect whether or not speech is present in an analysis frame. If frame i is such a frame, then $Q^{(i)}$ is an estimate of the noise spectrum. Since the noise spectrum contains information about part of what is likely to be present in a later frame j that contains speech plus noise, it follows that using $Q^{(i)}$ as a prior for frame j might result in more accurate estimation of the total spectrum in that frame, thus allowing more accurate compensation for the noise, say by subtraction of the noise spectrum. (On the other hand we might worry that this procedure would unduly enhance the noise component of the later estimate, thus further degrading the speech.) However, if the analyses of frames i and j are based on the same set of autocorrelation lags, prior washout occurs, and the use of $Q^{(i)}$ as a prior for frame j has no effect whatever on the result $Q^{(j)}$ of the analysis in frame j.

Although the same property holds for multi-signal MCESA, a combination of single-signal and multi-signal MCESA can be used to avoid prior washout and exploit the results of analyzing frames containing noise only. In particular, during a frame when speech is absent, obtain an estimated noise spectrum by a single-signal analysis. Use this spectrum as a prior noise estimate, together with some other appropriate spectrum as a prior speech estimate, for a multi-signal analysis in later frames. A procedure of this sort is illustrated in section IV.

The reason that prior washout does not occur in this case is that the initial computation of the estimated noise spectrum uses constraints on noise autocorrelations values, while the subsequent computations use constraints on total autocorrelations; thus different sets of functions are being constrained. In fact, let P_N be the prior used in obtaining the initial estimated noise spectrum $Q_N^{(1)}$ by single-signal MCESA. Then $Q_N^{(1)}$ has components at frequency f_k of the form

$$Q_{Nk}^{(1)} = \frac{1}{1/P_{Nk} + \sum_r \beta_r c_{rk}} .$$

If $Q_N^{(1)}$ is used as a noise prior in later computations, and a spectrum P_S is used as a speech prior, the resulting noise and speech posteriors $Q_N^{(2)}$ and $Q_S^{(2)}$ have the form

$$Q_{Nk}^{(2)} = \frac{1}{1/P_{Nk} + \sum_r \beta_r c_{rk} + \sum_r \beta'_r c_{rk}}, \tag{21}$$

$$Q_{Sk}^{(2)} = \frac{1}{1/P_{Sk} + \sum_r \beta'_r c_{rk}} . \tag{22}$$

If P_N were used in place of $Q_N^{(1)}$ in the later computations, the resulting posteriors would have the form

$$Q_{Nk}^{(2)} = \frac{1}{1/P_{Nk} + \sum_r \beta_r^* c_{rk}} , \tag{23}$$

$$Q_{Sk}^{(2)} = \frac{1}{1/P_{Sk} + \sum_r \beta_r^* c_{rk}} . \tag{24}$$

Now, for linearly independent constraints, (21) and (23) are compatible only if $\beta^*_r = \beta_r + \beta'_r$ holds, and (22) and (24) are compatible only if $\beta^*_r = \beta'_r$ holds. Thus the analog of prior washout will not in general occur here unless $\beta_r = 0$ holds -- that is, unless $Q_N^{(1)} = P_N$.

IV. EXAMPLES

In this section we present two numerical examples; in each, a given set of data is analyzed both by multi-signal MCESA and by either single-signal MCESA or a conventional MESA method. In the first example, autocorrelations at a small number of equally spaced lags are computed from the sum of a pair of assumed "true" spectra, and single- and multi-signal MCESA estimates are obtained from them. In the second, autocorrelations are estimated from sums of speech-signal and noise samples, and spectral estimates are obtained by MESA and multi-signal MCESA.

The assumed original spectra for the first example are a pair S_B and S_S, which we think of as a known "background" component and an unknown "signal" component of the total spectrum. For numerical purposes we use the spectral powers S_{Bk} and S_{Sk} at a hundred equally spaced frequencies $f_k = \pm.005, \pm.015, \ldots, \pm.495$ between $-.5$ and $+.5$ (the Nyquist band: we take the spacing between autocorrelation lags to be unity). The background consists of an approximation to white noise plus a peak corresponding to a sinusoid at frequency .215:

$$S_{Bk} = \begin{cases} 1.05, & f_k = \pm .215 \\ .05, & \text{otherwise} . \end{cases}$$

The signal term consists of a nearby, similar peak at frequency .165:

$$S_{Sk} = \begin{cases} 1, & f_k = \pm .165 \\ 0, & \text{otherwise} . \end{cases}$$

Thus the total assumed spectrum $S_B + S_S$ is as shown (for positive frequencies) in figure 1. Here are the corresponding autocorrelations R_r at six lags $t_r = 0, 1, \ldots , 5$:

t_r	0	1	2	3	4	5
R_r	9.0000	1.4544	-2.7732	-3.2248	0.2032	2.6900

For the multi-signal calculation, we use a pair of prior spectral estimates P_B and P_S. Since we are assuming prior knowledge of the background spectral component S_B, we simply take $P_B = S_B$ as shown in figure 2. To reflect prior ignorance of the signal component S_S, we take P_S to be uniform as in figure 3; for this example we have somewhat arbitrarily normalized P_S to have the same total power as P_B. For the single-signal calculation, we use $P = P_B + P_S$ as the prior spectral estimate.

Figure 4 shows the result of the single-signal analysis -- the MCESA posterior estimate Q obtained from the prior estimate P and autocorrelations R_r. Corresponding to the "known" peak at frequency .215 (which was included in the prior) there is a sharp peak in the posterior at that frequency; corresponding to the "unknown" peak at frequency .165 there is a maximum at

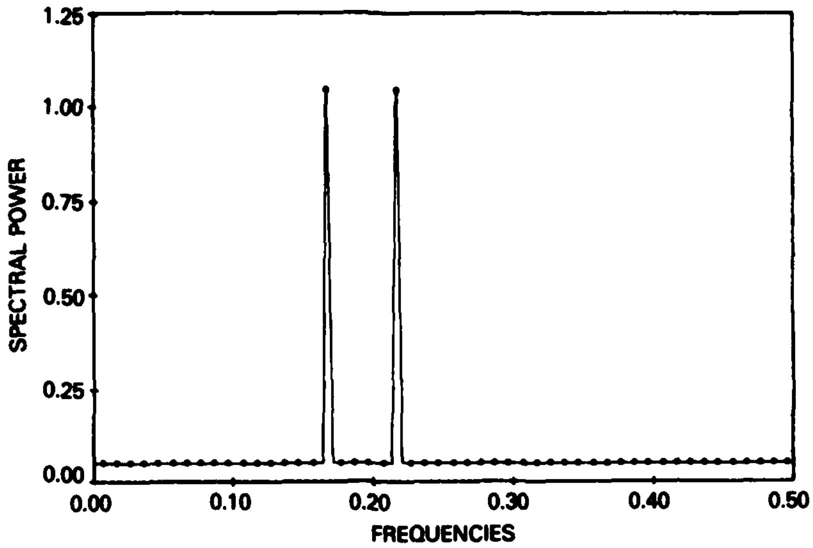


Fig. 1 -- Total assumed original spectrum (first example)

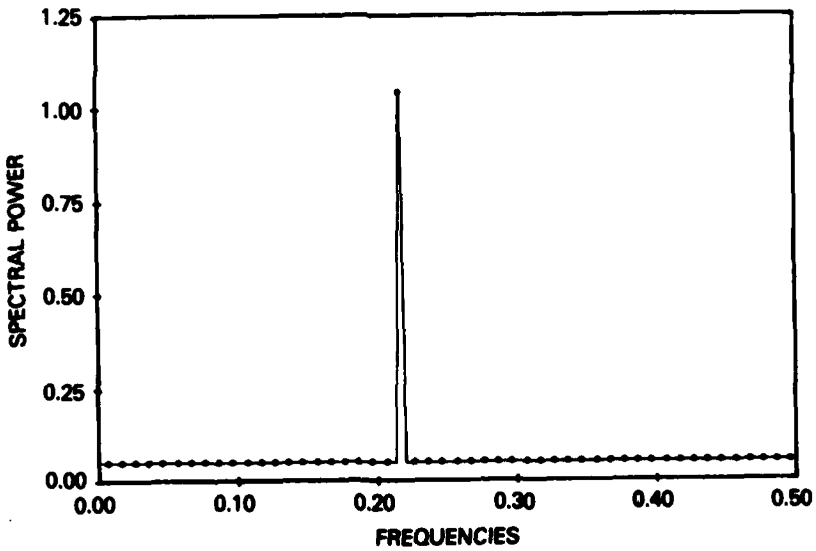


Fig. 2 -- Prior estimate of background spectrum: assumed original background spectrum (first example)

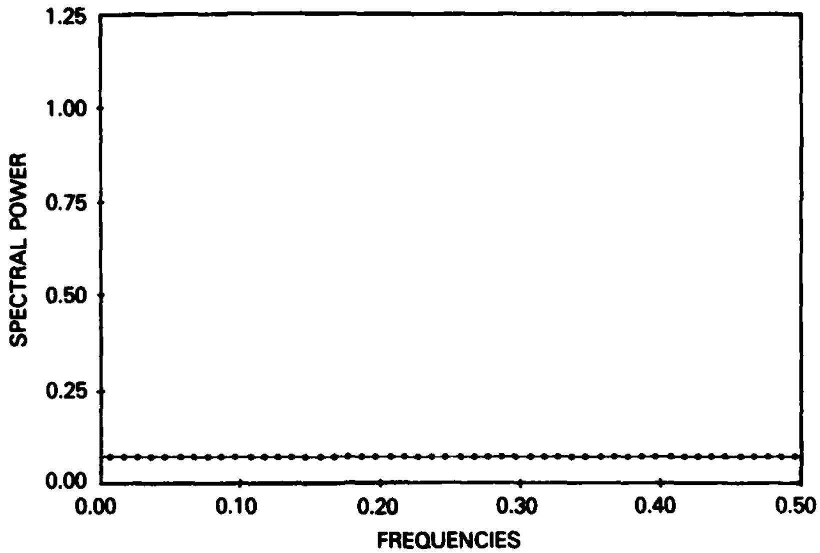


Fig. 3 -- Prior estimate of signal spectrum: uniform (first example)

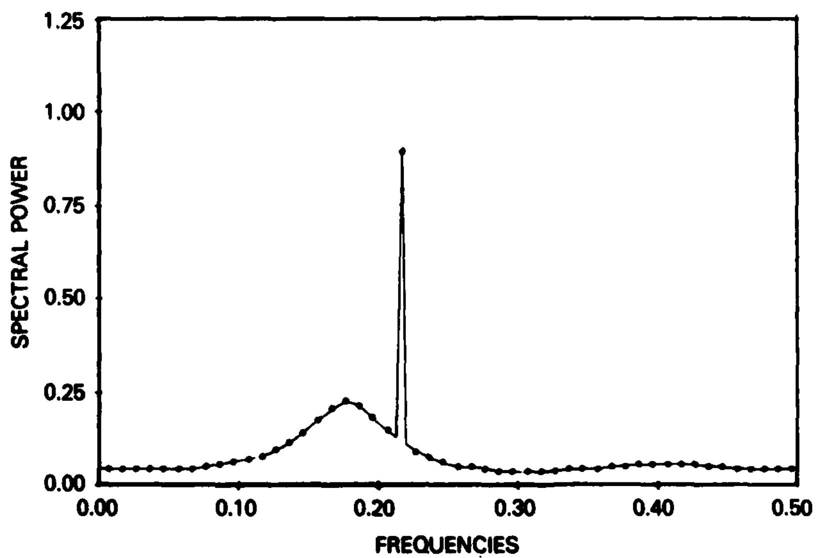


Fig. 4 -- Single-signal MCESA posterior estimate of total spectrum (first example)

approximately that frequency that is broader than the first, but resolvable from it.

The same original spectra S_B and S_S and the same autocorrelations R_r were used in an example in [1]. There a MESA and an MCESA spectral estimate were compared (see figures 5 and 6 in [1]). The MESA estimate failed to resolve the two peaks and showed a single maximum at about the mid-frequency. The MCESA estimate was based on P_B instead of $P_B + P_S$ as a prior; the result differed from figure 4, but was qualitatively similar. In both cases, the MCESA estimate implies the presence of the signal at frequency .165, but does not provide a numerical estimate of the signal. Such an estimate is provided by multi-signal MCESA.

The two individual posterior estimates Q_B and Q_S from the multi-signal analysis are shown in figures 5 and 6. The sharp peak at frequency .215 is seen to be correctly assigned entirely to the background posterior Q_B -- unsurprisingly, since it was present in the background prior but not the signal prior. The broader maximum corresponding to the original peak at frequency .165 is present in the signal posterior Q_S and is also present, though less prominent, in Q_B. To understand why, qualitatively, consider that the autocorrelations depend only on the total spectrum; the autocorrelation constraints can equally well be satisfied by allocating spectral power near frequency .165 to Q_B, or to Q_S. By the discussion in section III, the relative magnitudes of the posteriors at each frequency depend on the relative magnitudes of the priors. Both P_B and P_S are flat near frequency .165, and because of the normalization chosen, P_S is somewhat greater there. Consequently, the broad maximum in Q_S is somewhat greater than that in Q_B.

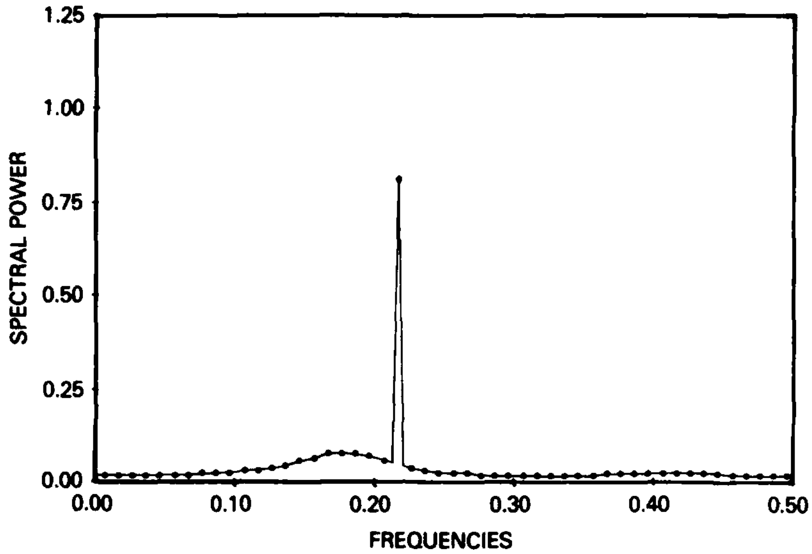


Fig. 5 -- Multi-signal MCESA posterior estimate of background spectrum (first example)

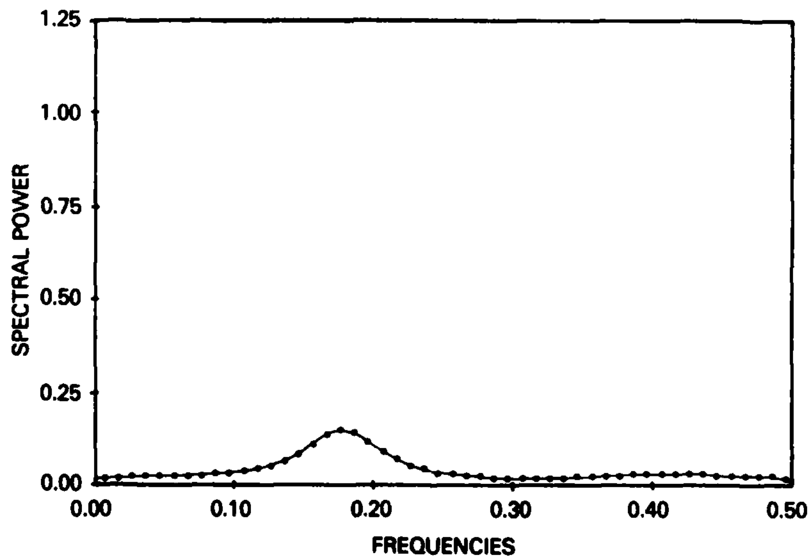


Fig. 6 -- Multi-signal MCESA posterior estimate of signal spectrum (first example)

The second example is based on time-domain samples of voiced speech and noise. The speech comprises a portion of an English sentence spoken by a male speaker and includes the first word, "Sue," of the sentence together with silent segments before and after it. The noise consists of a segment of helicopter noise equal in duration to the speech. These were separately filtered, sampled, and digitized at 8000 samples per second. The speech and noise data were then added sample by sample, resulting in samples of noisy speech. These samples were segmented into analysis frames of 180 samples, and 11 autocorrelations R_r, $r = 0, 1, \ldots, 10$ were estimated for each frame by the formula

$$R_r = \frac{1}{180} \sum_{j=1}^{180-r} s_j s_{j+r} ,$$

where s_j is the $j^{\underline{th}}$ sample in the frame. This is a biased estimate but guarantees positive-definiteness. No additional windowing or filtering was used.

The last frame before the actual beginning of the word was selected; this frame of "noisy speech" thus consisted entirely of noise. From the autocorrelation estimates for this frame, a conventional MESA (i.e. uniform-prior MCESA) spectral estimate was computed for use as a prior estimate of the noise spectrum in subsequent frames. A uniform spectrum was used as a prior estimate for the speech spectrum in the subsequent frames. These two priors are shown in figure 7. Much of the noise power is concentrated in a peak near 2780 Hz.

From the two priors and the autocorrelation estimates, multi-signal MCESA estimates of the speech and noise spectra were computed for later frames. From the autocorrelation estimates, MESA (LPC) spectral estimates were

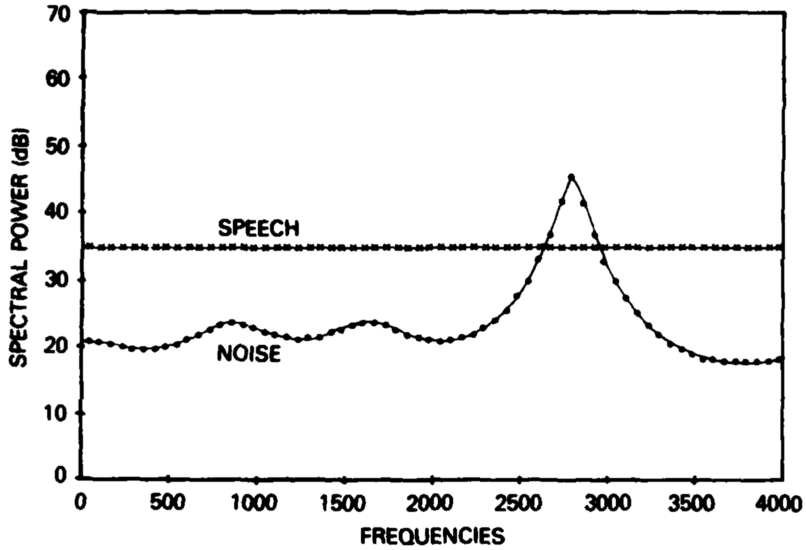


Fig. 7 -- Prior estimates of speech and noise spectra (second example)

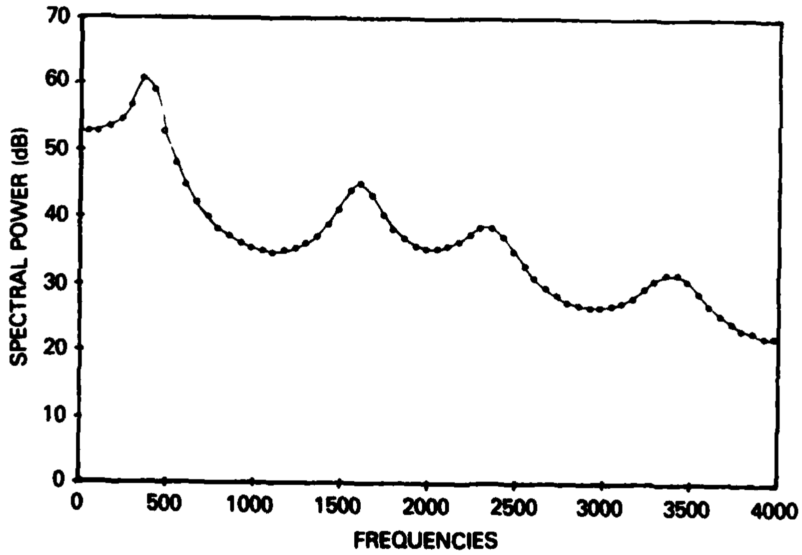


Fig. 8 -- MESA estimate of speech spectrum from noise-free data (second example)

computed for the noisy speech. We present the results for a selected frame of voiced speech--the second of seven that span the vowel /u/. For comparison with these results, we present in figure 8 a MESA estimate of the uncorrupted speech. This was computed exactly like the MESA estimate for the noisy speech except that the R_r were estimated from the speech samples only, not from the sums of speech and noise samples.

The MESA estimate for the noisy speech is shown in figure 9. This spectrum agrees rather well with the noise-free estimate in the band from 0 up to about 2000 Hz, which includes the first two formants. Above 2000 Hz, however, there is only a single maximum; the third and fourth formants have merged with the peak in the noise spectrum to form a single peak at about 2690 Hz.

We subtracted the noise prior (figure 7) from this result (figure 9). The difference, shown in figure 10, represents an attempt to estimate the speech spectrum by a MESA analysis and spectral subtraction. The subtracted MESA spectrum is fairly close to the unsubtracted MESA spectrum except in the neighborhood of the noise peak at 2780 Hz. Near that frequency, the subtraction so far overcompensates that the difference actually assumes rather large negative values. (Absolute values are plotted in the figure.)

The multi-signal MCESA posteriors are shown in figures 11 and 12; figure 11 is the speech, and figure 12 is the noise. Figure 12 shows a maximum near 2440 Hz, about 130 Hz higher than the third formant, and a suggestion of the fourth formant is discernible. Except for frequencies near the noise peak, the multi-signal speech spectrum (figure 11) and the subtracted MESA result (figure 10) are quite close, the multi-signal result being usually the closer of the two to the estimate based on noise-free data (figure 8). Near 2780 Hz,

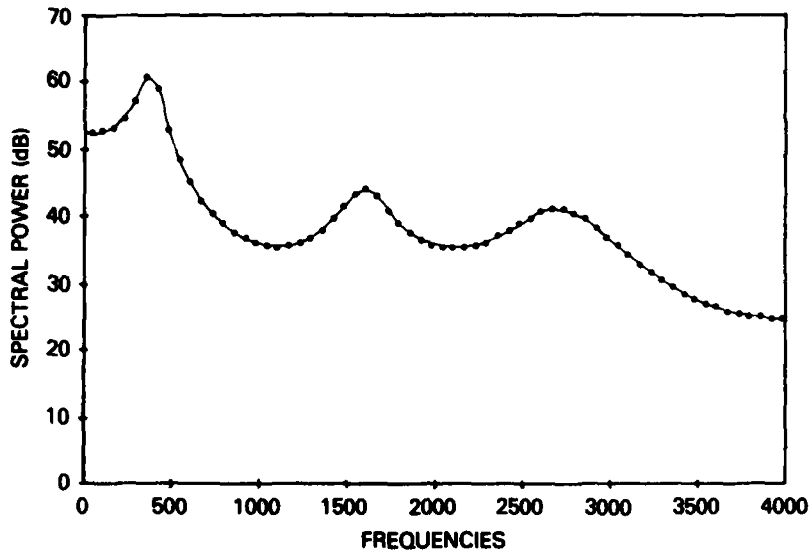


Fig. 9 — MESA estimate of total spectrum (second example)

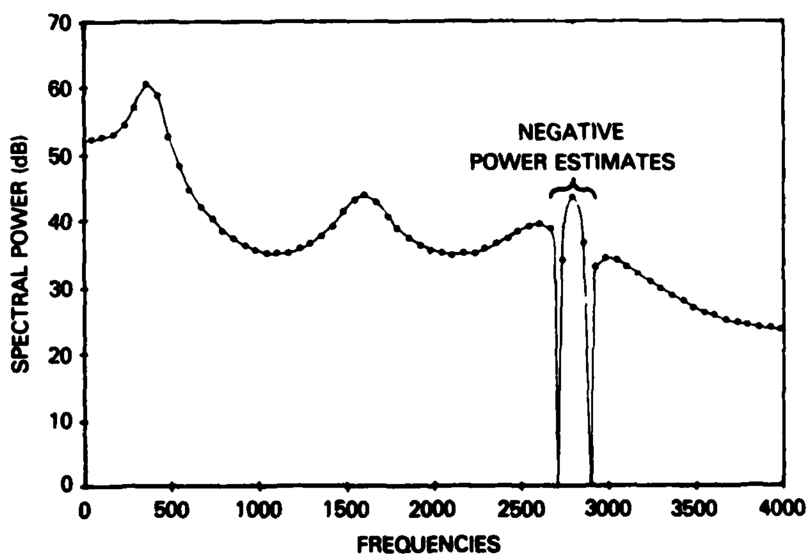


Fig. 10 — Result of subtracting noise prior (see figure 7) from spectrum in figure 9 (second example)

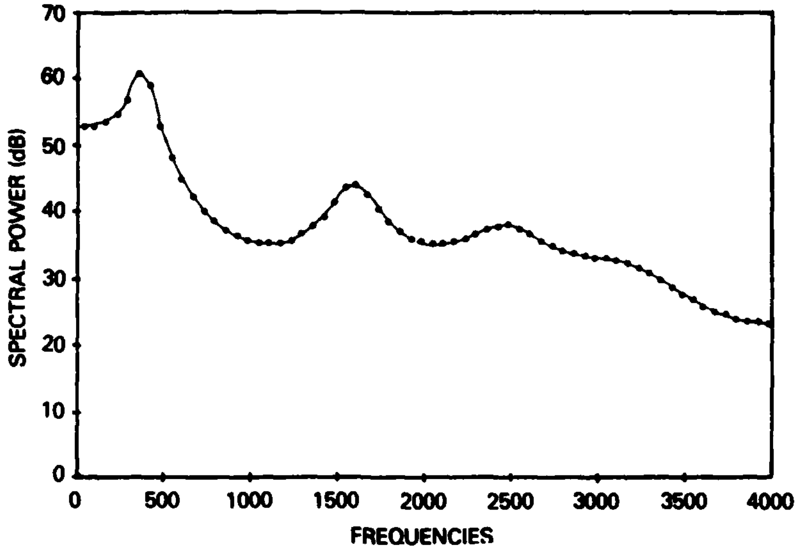


Fig. 11 — Multi-signal MCESA posterior estimate of speech spectrum (second example)

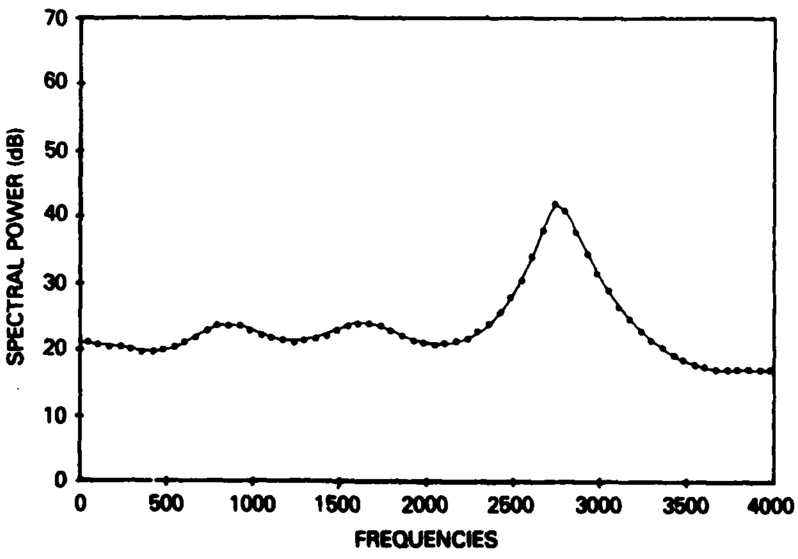


Fig. 12 — Multi-signal MCESA posterior estimate of noise spectrum (second example)

the multi-signal result is substantially closer, and where the subtracted MESA becomes negative, the multi-signal estimate takes only physically meaningful positive values. Both methods underestimate the total power near 2780 Hz (cf. figure 12); however, the multi-signal method apportions the total between speech and noise in a somewhat reasonable way, whereas the other does not.

V. DISCUSSION AND CONCLUSIONS

Multi-signal MCESA is a new spectrum-estimation method based on a provably optimal information-theoretic inductive-inference procedure. When separate prior estimates are available for the power spectra of two or more processes, and new information is obtained in the form of values of the autocorrelation function of their sum, the method yields separate posterior estimates. One suggested application is separating the spectrum of a signal from that of *additive noise*. By incorporating prior estimates for both signal and noise spectra, the multi-signal method offers considerable scope and flexibility for tailoring an estimator to the characteristics of a signal or noise.

In the second example in section IV we contrasted this method with a more ad-hoc method for taking a prior noise estimate into account -- estimate the sum of signal and noise spectra from autocorrelations and then subtract the prior noise estimate. The latter method seems to imply an unwarranted absolute commitment to the noise-spectrum estimate: adjustments to the signal-spectrum estimate are made solely responsible for fitting the autocorrelation of the sum to measured values. The multi-signal method, by contrast, adjusts both noise and signal estimates in fitting the autocorrelation of the sum. We saw that the multi-signal method could thereby avoid nonphysical (negative) estimates that can result from spectral subtraction. Of course, whether or not the multi-signal method can improve speech quality must be determined by systematic experiments involving speech synthesis and intelligibility testing. We hope to do such experiments in the future.

In the same example, a prominent noise peak was present in the sum spectrum. Most of the power in it was properly attributed to the noise spectrum in the posteriors, but substantial leakage a few db lower into the signal (speech) spectrum occurred. The relative apportionment of the power in that peak between the signal and noise posteriors would be substantially altered by a change in the level of the uniform spectrum that was used as the speech prior. This is in contrast to single-signal MCESA, where all uniform priors give equivalent results (as long as one of the constrained autocorrelations is the total power). How best to choose the level of this uniform prior relative to the noise prior is a question not yet answered. Indeed, since the signal is known to be speech, it would undoubtedly be beneficial to replace the uniform signal prior with one tailored to the characteristics of speech. How best to do this tailoring is another unanswered question. In short, there is much to be learned about how to choose the prior estimates to reflect our prior knowledge of signals and noise in practical situations.

VI. ACKNOWLEDGMENTS

We thank George Kang for supplying the speech and noise data on which the second numerical example was based, and for a number of helpful discussions.

REFERENCES

[1] J.E. Shore, "Minimum Cross-Entropy Spectral Analysis," IEEE Trans. Acoustics, Speech, & Signal Proc. to appear.

[2] J. P. Burg, "Maximum Entropy Spectral Analysis," presented at the 37th Annual Meeting Soc. of Exploration Geophysicists, Oklahoma City, Okla., 1967.

[3] J. Burg, "Maximum Entropy Spectral Analysis," Ph.D. Dissertation, Stanford University, 1975 (University Microfilms No. 75-25,499).

[4] W.M. Elsasser, "On Quantum Measurements and the Role of the Uncertainty Relations in Statistical Mechanics," Phys. Rev., vol. 52, pp. 987-999, Nov. 1937.

[5] E. T. Jaynes, "Prior Probabilities," IEEE Trans. Systems Science and Cybernetics, vol. SSC-4, pp. 227-241, 1968.

[6] S. Kullback, Information Theory and Statistics, Wiley, New York, 1959; Dover, New York, 1969.

[7] J.E. Shore and R.W. Johnson, "Axiomatic Derivation of the Principle of Maximum Entropy and the Principle of Minimum Cross-Entropy," IEEE Trans. Inf. Theory, vol. IT-26, no. 1, pp. 26-37, Jan. 1980.

[8] I. J. Good, "Maximum Entropy for Hypothesis Formulation, Especially for Multidimensional Contingency Tables," Annals Math. Stat., vol. 34, pp. 911-934, 1963.

[9] A. Hobson and B. Cheng, "A Comparison of the Shannon and Kullback Information Measures," J. Stat. Phys., vol. 7, No. 4, pp. 301-310, 1973.

[10] I. Csiszár, "I-Divergence Geometry of Probability Distributions and Minimization Problems," Ann. Prob., vol. 3, pp. 146-158, 1975.

[11] M.S. Pinsker, Information and Information Stability of Random Variables and Processes, Holden-Day, San Francisco, 1964.

[12] F. Itakura and S. Saito, "Analysis Synthesis Telephone Based upon the Maximum Likelihood Method," Reports of the 6th Int. Cong. Acoustics, Y. Yonsai, ed., Tokyo, 1968.

[13] R. Johnson, "Axiomatic Characterization of the Directed Divergences and Their Linear Combinations," IEEE Trans. Inform. Theory, vol. IT-25, no. 6, pp. 709-716, Nov. 1979.

[14] J.E. Shore and R.W. Johnson, "Properties of Cross-Entropy Minimization," IEEE Trans. Inform. Theory, to appear July 1981.

[15] L. Geronimus, Orthogonal Polynomials, Consultants bureau, New York, 1961.

[16] R.M. Gray, A. Buzo, A.H. Gray, Jr., and Y. Matsuyama, "Distortion Measures for Speech Processing," IEEE Trans. Acoustics, Speech, and Signal Processing, vol. ASSP-28, pp. 367-376, Aug. 1980.

[17] J.D. Markel and A.H. Gray, Jr., Linear Prediction of Speech, Springer-Verlag, New York, 1976.

[18] Y. Yaglom, An Introduction to the Theory of Stationary Random Functions, Prentice-Hall, New York, 1962.

[19] J.E. Shore, "Minimum Cross-Entropy Spectral Analysis," NRL Memorandum Rept. 3921, Naval Research Laboratory, Washington, D.C. 20375, January, 1979.

DATE

ILMED

-8